

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Wei, et al.
- (ii) TITLE OF INVENTION: Transforming Growth Factor Alpha HII
- (iii) NUMBER OF SEQUENCES: 14
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Human Genome Sciences, Inc.
 - (B) STREET: 9410 Key West Ave
 - (C) CITY: Rockville
 - (D) STATE: MD
 - (E) COUNTRY: US
 - (F) ZIP: 20850
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 09/227,853
 - (B) FILING DATE: 11-JAN-1999
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: 08/930,564
 - (B) FILING DATE: 30-JAN-1998
- (viii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: WO PCT/US95/06386
 - (B) FILING DATE: 19-MAY-1995
- (ix) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Brookes, A. Anders
 - (B) REGISTRATION NUMBER: 36,373
 - (C) REFERENCE/DOCKET NUMBER: PF174USD1
- (x) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 301-309-8504
 - (B) TELEFAX: 301-309-8439

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1695 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS

(B) LOCATION: 323..1444

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CACTCGTCTG CCCCTGGACT CCCGTCTCCT CCTGTCCTCC GGCTTCCCAG AGCTCCCTCC	60
TTATGGCAGC AGCTTCCCCG GTCTCCGGCG CAGTTCTCAG CGGACGACCC TCTCGCTCCG	120
GGGCTGAGCC CAGTCCCTGG ATGTTGCTGA AACTCTCGAG ATCATGCGCG GGTTTGGCTG	180
CTGCTTCCCC GCCGGGTGCC ACTGCCACCG CCGCCGCCTC TGCTGCCGCC GTCCGCGGGA	240
TGCTCAGTAG CCCGCTGCCC GGCCCCCGCG ATCCTGTGTT CCTCGGAAGC CGTTTGCTGC	300
TGCAGAGTTG CACGAACTAG TC ATG GTG CTG TGG GAG TCC CCG CGG CAG TGC	352
Met Val Leu Trp Glu Ser Pro Arg Gln Cys	
1 5 10	
AGC AGC TGG ACA CTT TGC GAG GGC TTT TGC TGG CTG CTG CTG CTG CCC	400
Ser Ser Trp Thr Leu Cys Glu Gly Phe Cys Trp Leu Leu Leu Leu Pro	
15 20 25	
GTC ATG CTA CTC ATC GTA GCC CGC CCG GTG AAG CTC GCT GCT TTC CCT	448
Val Met Leu Leu Ile Val Ala Arg Pro Val Lys Leu Ala Ala Phe Pro	
30 35 40	
ACC TCC TTA AGT GAC TGC CAA ACG CCC ACC GGC TGG AAT TGC TCT GGT	496
Thr Ser Leu Ser Asp Cys Gln Thr Pro Thr Gly Trp Asn Cys Ser Gly	
45 50 55	
TAT GAT GAC AGA GAA AAT GAT CTC TTC CTC TGT GAC ACC AAC ACC TGT	544
Tyr Asp Asp Arg Glu Asn Asp Leu Phe Leu Cys Asp Thr Asn Thr Cys	
60 65 70	
AAA TTT GAT GGG GAA TGT TTA AGA ATT GGA GAC ACT GTG ACT TGC GTC	592
Lys Phe Asp Gly Glu Cys Leu Arg Ile Gly Asp Thr Val Thr Cys Val	
75 80 85 90	
TGT CAG TTC AAG TGC AAC AAT GAC TAT GTG CCT GTG TGT GGC TCC AAT	640
Cys Gln Phe Lys Cys Asn Asn Asp Tyr Val Pro Val Cys Gly Ser Asn	
95 100 105	
GGG GAG AGC TAC CAG AAT GAG TGT TAC CTG CGA CAG GCT GCA TGC AAA	688
Gly Glu Ser Tyr Gln Asn Glu Cys Tyr Leu Arg Gln Ala Ala Cys Lys	
110 115 120	
CAG CAG AGT GAG ATA CTT GTG GTG TCA GAA GGA TCA TGT GCC ACA GAT	736
Gln Gln Ser Glu Ile Leu Val Val Ser Glu Gly Ser Cys Ala Thr Asp	
125 130 135	
GCA GGA TCA GGA TCT GGA GAT GGA GTC CAT GAA GGC TCT GGA GAA ACT	784
Ala Gly Ser Gly Ser Gly Asp Gly Val His Glu Gly Ser Gly Glu Thr	
140 145 150	
AGT CAA AAG GAG ACA TCC ACC TGT GAT ATT TGC CAG TTT GGT GCA GAA	832
Ser Gln Lys Glu Thr Ser Thr Cys Asp Ile Cys Gln Phe Gly Ala Glu	
155 160 165 170	
TGT GAC GAA GAT GCC GAG GAT GTC TGG TGT GTG TGT AAT ATT GAC TGT	880

Cys	Asp	Glu	Asp	Ala	Glu	Asp	Val	Trp	Cys	Val	Cys	Asn	Ile	Asp	Cys	
				175					180					185		
TCT	CAA	ACC	AAC	TTC	AAT	CCC	CTC	TGC	GCT	TCT	GAT	GGG	AAA	TCT	TAT	928
Ser	Gln	Thr	Asn	Phe	Asn	Pro	Leu	Cys	Ala	Ser	Asp	Gly	Lys	Ser	Tyr	
			190					195					200			
GAT	AAT	GCA	TGC	CAA	ATC	AAA	GAA	GCA	TCG	TGT	CAG	AAA	CAG	GAG	AAA	976
Asp	Asn	Ala	Cys	Gln	Ile	Lys	Glu	Ala	Ser	Cys	Gln	Lys	Gln	Glu	Lys	
		205					210					215				
ATT	GAA	GTC	ATG	TCT	TTG	GGT	CGA	TGT	CAA	GAT	AAC	ACA	ACT	ACA	ACT	1024
Ile	Glu	Val	Met	Ser	Leu	Gly	Arg	Cys	Gln	Asp	Asn	Thr	Thr	Thr	Thr	
	220					225					230					
ACT	AAG	TCT	GAA	GAT	GGG	CAT	TAT	GCA	AGA	ACA	GAT	TAT	GCA	GAG	AAT	1072
Thr	Lys	Ser	Glu	Asp	Gly	His	Tyr	Ala	Arg	Thr	Asp	Tyr	Ala	Glu	Asn	
235					240					245					250	
GCT	AAC	AAA	TTA	GAA	GAA	AGT	GCC	AGA	GAA	CAC	CAC	ATA	CCT	TGT	CCG	1120
Ala	Asn	Lys	Leu	Glu	Glu	Ser	Ala	Arg	Glu	His	His	Ile	Pro	Cys	Pro	
			255						260					265		
GAA	CAT	TAC	AAT	GGC	TTC	TGC	ATG	CAT	GGG	AAG	TGT	GAG	CAT	TCT	ATC	1168
Glu	His	Tyr	Asn	Gly	Phe	Cys	Met	His	Gly	Lys	Cys	Glu	His	Ser	Ile	
			270					275					280			
AAT	ATG	CAG	GAG	CCA	TCT	TGC	AGG	TGT	GAT	GCT	GGT	TAT	ACT	GGA	CAA	1216
Asn	Met	Gln	Glu	Pro	Ser	Cys	Arg	Cys	Asp	Ala	Gly	Tyr	Thr	Gly	Gln	
		285					290					295				
CAC	TGT	GAA	AAA	AAG	GAC	TAC	AGT	GTT	CTA	TAC	GTT	GTT	CCC	GGT	CCT	1264
His	Cys	Glu	Lys	Lys	Asp	Tyr	Ser	Val	Leu	Tyr	Val	Val	Pro	Gly	Pro	
	300					305					310					
GTA	CGA	TTT	CAG	TAT	GTC	TTA	ATC	GCA	GCT	GTG	ATT	GGA	ACA	ATT	CAG	1312
Val	Arg	Phe	Gln	Tyr	Val	Leu	Ile	Ala	Ala	Val	Ile	Gly	Thr	Ile	Gln	
315					320					325					330	
ATT	GCT	GTC	ATC	TGT	GTG	GTG	GTC	CTC	TGC	ATC	ACA	AGG	AAA	TGC	CCC	1360
Ile	Ala	Val	Ile	Cys	Val	Val	Val	Leu	Cys	Ile	Thr	Arg	Lys	Cys	Pro	
				335					340					345		
AGA	AGC	AAC	AGA	ATT	CAC	AGA	CAG	AAG	CAA	AAT	ACA	GGG	CAC	TAC	AGT	1408
Arg	Ser	Asn	Arg	Ile	His	Arg	Gln	Lys	Gln	Asn	Thr	Gly	His	Tyr	Ser	
			350					355					360			
TCG	GAC	AAT	ACA	ACA	AGA	GCG	TCC	ACG	AGG	TTA	ATC	TAAAGGGAGC				1454
Ser	Asp	Asn	Thr	Thr	Arg	Ala	Ser	Thr	Arg	Leu	Ile					
		365					370									
ATGTTTCACA	GTGGCTGGAC	TACCGAGAGC	TTGGACTACA	CAATACAGTA	TTATAGACAA											1514
AAGAATAAGA	CAAGAGATCT	ACACATGTTG	CCTTGCAATTT	GTGGTAATCT	ACACCAATGA											1574
AAACATGTAC	TACAGCTATA	TTTGATTATG	TATGGATATA	TTTGAAATAG	TATACATTGT											1634
CTTGATGTTT	TTTCTGTAAT	GTAAATAAAC	TATTTATATC	ACACAAAAAA	AAAAAAAAAA											1694
A																1695

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 374 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Val	Leu	Trp	Glu	Ser	Pro	Arg	Gln	Cys	Ser	Ser	Trp	Thr	Leu	Cys
1				5					10					15	
Glu	Gly	Phe	Cys	Trp	Leu	Leu	Leu	Leu	Pro	Val	Met	Leu	Leu	Ile	Val
			20					25						30	
Ala	Arg	Pro	Val	Lys	Leu	Ala	Ala	Phe	Pro	Thr	Ser	Leu	Ser	Asp	Cys
			35				40						45		
Gln	Thr	Pro	Thr	Gly	Trp	Asn	Cys	Ser	Gly	Tyr	Asp	Asp	Arg	Glu	Asn
	50					55					60				
Asp	Leu	Phe	Leu	Cys	Asp	Thr	Asn	Thr	Cys	Lys	Phe	Asp	Gly	Glu	Cys
65					70					75					80
Leu	Arg	Ile	Gly	Asp	Thr	Val	Thr	Cys	Val	Cys	Gln	Phe	Lys	Cys	Asn
				85					90					95	
Asn	Asp	Tyr	Val	Pro	Val	Cys	Gly	Ser	Asn	Gly	Glu	Ser	Tyr	Gln	Asn
			100					105					110		
Glu	Cys	Tyr	Leu	Arg	Gln	Ala	Ala	Cys	Lys	Gln	Gln	Ser	Glu	Ile	Leu
		115					120					125			
Val	Val	Ser	Glu	Gly	Ser	Cys	Ala	Thr	Asp	Ala	Gly	Ser	Gly	Ser	Gly
	130					135					140				
Asp	Gly	Val	His	Glu	Gly	Ser	Gly	Glu	Thr	Ser	Gln	Lys	Glu	Thr	Ser
145					150				155						160
Thr	Cys	Asp	Ile	Cys	Gln	Phe	Gly	Ala	Glu	Cys	Asp	Glu	Asp	Ala	Glu
				165					170					175	
Asp	Val	Trp	Cys	Val	Cys	Asn	Ile	Asp	Cys	Ser	Gln	Thr	Asn	Phe	Asn
			180					185					190		
Pro	Leu	Cys	Ala	Ser	Asp	Gly	Lys	Ser	Tyr	Asp	Asn	Ala	Cys	Gln	Ile
		195					200					205			
Lys	Glu	Ala	Ser	Cys	Gln	Lys	Gln	Glu	Lys	Ile	Glu	Val	Met	Ser	Leu
	210					215					220				
Gly	Arg	Cys	Gln	Asp	Asn	Thr	Thr	Thr	Thr	Thr	Lys	Ser	Glu	Asp	Gly
225					230					235					240
His	Tyr	Ala	Arg	Thr	Asp	Tyr	Ala	Glu	Asn	Ala	Asn	Lys	Leu	Glu	Glu
				245					250					255	
Ser	Ala	Arg	Glu	His	His	Ile	Pro	Cys	Pro	Glu	His	Tyr	Asn	Gly	Phe

	260		265		270										
Cys	Met	His	Gly	Lys	Cys	Glu	His	Ser	Ile	Asn	Met	Gln	Glu	Pro	Ser
		275					280					285			
Cys	Arg	Cys	Asp	Ala	Gly	Tyr	Thr	Gly	Gln	His	Cys	Glu	Lys	Lys	Asp
	290					295					300				
Tyr	Ser	Val	Leu	Tyr	Val	Val	Pro	Gly	Pro	Val	Arg	Phe	Gln	Tyr	Val
305					310					315					320
Leu	Ile	Ala	Ala	Val	Ile	Gly	Thr	Ile	Gln	Ile	Ala	Val	Ile	Cys	Val
				325					330					335	
Val	Val	Leu	Cys	Ile	Thr	Arg	Lys	Cys	Pro	Arg	Ser	Asn	Arg	Ile	His
			340					345					350		
Arg	Gln	Lys	Gln	Asn	Thr	Gly	His	Tyr	Ser	Ser	Asp	Asn	Thr	Thr	Arg
		355					360					365			
Ala	Ser	Thr	Arg	Leu	Ile										
	370														

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCCGGATCCG CACGAGACAT ACCTTGTCGG

30

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGGAAGCTTT TAATACTGAA ATCGTACAGG AC

32

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CGCGGATCCG CCATCATGGT GCTGTGGGAG TCC

33

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GCGTCTAGAC TAGTATAGAA CACTGTAGTC C

31

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCGTCTAGAC TAGTATAGAA CACTGTAGTC C

31

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCGTCTAGAC TAGTATAGAA CACTGTAGTC C

31

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CGCGGATCCA GAACACCACA TACCTTGTCC G

31

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GCGTCTAGAC TAGTATAGAA CACTGTAGTC C

31

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CGCGGATCCG CCATCATGGT GCTGTGGGAG TCC

33

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GCGCTCGAGG TATAGAACAC TGTAGTCC

28

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 177 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Met Asp Arg Ala Ala Arg Cys Ser Gly Ala Ser Ser Leu Pro Leu Leu
1 5 10 15

Leu Ala Leu Ala Leu Gly Leu Val Ile Leu His Cys Val Val Ala Asp
20 25 30

Gly Asn Ser Thr Arg Ser Pro Glu Thr Asn Gly Leu Leu Cys Gly Asp
35 40 45

Pro Glu Glu Asn Cys Ala Ala Thr Thr Thr Gln Ser Lys Arg Lys Gly
50 55 60

His Phe Ser Arg Cys Pro Lys Gln Tyr Lys His Tyr Cys Ile Lys Gly
65 70 75 80

Arg Cys Arg Phe Val Val Ala Glu Gln Thr Pro Ser Cys Val Cys Asp
85 90 95

Glu Gly Tyr Ile Gly Ala Arg Cys Glu Arg Val Asp Leu Phe Tyr Leu
100 105 110

Arg Gly Asp Arg Gly Gln Ile Leu Val Ile Leu Ile Ala Val Met Val
115 120 125

Val Phe Ile Ile Leu Val Ile Gly Val Cys Thr Cys Cys His Pro Leu
130 135 140

Arg Lys Arg Arg Lys Arg Lys Lys Lys Glu Glu Glu Met Glu Thr Leu
145 150 155 160

Gly Lys Asp Ile Thr Pro Ile Asn Glu Asp Ile Glu Glu Thr Asn Ile
165 170 175

Ala

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 159 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Met	Val	Pro	Ser	Ala	Gly	Gln	Leu	Ala	Leu	Phe	Ala	Leu	Gly	Ile	Val	
1				5				10						15		
Leu	Ala	Ala	Cys	Gln	Ala	Leu	Glu	Asn	Ser	Thr	Ser	Pro	Leu	Ser	Ala	
			20					25					30			
Asp	Pro	Pro	Val	Ala	Ala	Ala	Val	Val	Ser	His	Phe	Asn	Asp	Cys	Pro	
		35					40					45				
Asp	Ser	His	Thr	Phe	Cys	Phe	His	Gly	Thr	Cys	Arg	Phe	Leu	Val	Gln	
	50					55					60					
Glu	Asp	Lys	Pro	Ala	Cys	Val	Cys	His	Ser	Gly	Tyr	Val	Gly	Ala	Arg	
65					70					75					80	
Cys	Glu	His	Ala	Asp	Leu	Leu	Ala	Val	Val	Ala	Ala	Ser	Gln	Lys	Lys	
				85					90					95		
Gln	Ala	Ile	Thr	Ala	Leu	Val	Val	Val	Ser	Ile	Val	Ala	Leu	Ala	Val	
			100					105					110			
Leu	Ile	Ile	Thr	Cys	Val	Leu	Ile	His	Cys	Cys	Ala	Val	Arg	Lys	His	
	115						120					125				
Cys	Glu	Trp	Cys	Arg	Ala	Leu	Ile	Cys	Arg	His	Glu	Lys	Pro	Ser	Ala	
	130					135					140					
Leu	Leu	Lys	Gly	Arg	Thr	Ala	Cys	Cys	His	Ser	Glu	Thr	Val	Val		
145					150					155						